

Module for Machine Translating Relative and Non-Relative "That" from English into Arabic

FARAH ABBAS ABO AL TIMEN
Asst. Lecturer, Department of Translation
College of Arts, Al Mustansiriya University

Abstract:

This study provides a framework for Machine translating the relative and non-relative word "That" from English into Arabic. The choice of this subject in particular is based on the fact that the lack of accuracy in available machine translation software tools such as Google proved to be inefficient in translating the functions of this word. The mistranslation of these functions does not affect the meaning of this word only but the whole sentence meaning in different linguistic layers (syntactic and lexical).

Tests on 'that' were performed by other software tools (Google, Babylon, Golden Al-Wafi and Bing translators) to demonstrate the mistranslation of each software and set the suitable rules in order to be translated properly into Arabic.

A major design goal of this module is that it can be used as a stand-alone tool and can be very well integrated with a general machine translation system with an English-Arabic sentence. The findings of the study have validated the hypothesis; in addition, it has been found out that this module has a major design significance that is it can be integrated with a general machine translation system with an English-Arabic direction. The module is implemented in Visual Prolog.

Key words: Machine Translating, Relative and Non-Relative "That", English, Arabic

Introduction

Artificial intelligence is a branch of computer science and the intelligence of machines which aims at creating automatic tasks that human intelligence already has. Wilks (2009: 27) takes this branch to be 'the enterprise of causing automata to perform peculiarly human tasks'. Artificial intelligence explores the basic mechanisms of a language and mind by the attempt to model and simulate machine programs.

Machine translation is 'automate all or part of the process of translating from one human language to another' (Arnold et al, 1994: 1) or an 'automatic process that utilizes computer software to translate texts from one language to another' (Aljlal et al, 2004: 1). This definition stresses the fact that machine translation is not simply substituting words by other words, but like human translation, it should involve the mechanism of processing efficient translation.

'*That*' functions in English language as a relative and non-relative word. Such positions need to be fully understood and evaluated by machine. Since machine processing of structures is not similar to that of human mind, a more complicated syntactic analysis is required to identify the structure of a sentence to verify whether the structure in which '*that*' comes in is functioning as a relative or non-relative. In order to obtain an output which is close in accuracy to that produced by human beings, the computer needs to be provided with the appropriate procedures and it must go through three complicated barriers:

- The language barrier.
- The cross linguistic barrier.
- The translation barrier.

(Attia, 2002: 6)

Machine translation results are mostly inappropriate. The mistranslation of '*that*' (representing a relative clause and a

Demonstrative/ Deictic word) in all the software translators can cause damage to the meaning more than in any other clauses as is shown in the following table. The samples listed in Table 1 & 2 were translated by the most successful MT softwares currently in work: Google translator (2012)¹, Babylon (2012)², Golden Al-Wafi (2012)³ and Bing (2012)⁴.

Sentence		Translation by			
		Google	Babylon	Golden Al -Wafi	Bing
Relative 'That'	Today I met a woman that I used to work with.	اليوم التقيت امرأة <u>أن</u> كنت أعمل مع.	اليوم قابلت امرأة كنت أعمل مع.	اليوم اجتمعت امرأة أنا a كنت أعمل مع.	اليوم التقيت امرأة التي تستخدم للعمل مع.
	They selected an actor that is not fit for the part.	اختيارهم فاعل التي لا يصلح الجزء.	اختاروا فاعلا لا يليق.	اختاروا ممثلا الذي لم يلائم الجزء.	أنها تحديد فاعل لا تناسب الجزء.
	The two rivers that pass through Iraq are Tigris and Euphrates.	النهرين التي تمر عبر العراق دجلة والفرات.	بلاد الرافدين التي تمر عبر العراق والفرات. دجلة.	ان النهرين التي تعبران العراق هي دجلة و الفرات.	واثنين من الأنهار التي تمر عبر العراق هي دجلة والفرات.

Table (1) Samples of Software Translators Output of Relative 'That' Clauses Input

Non-Relative 'That'	Holly felt that her work there had been done.	شعرت هوللي الذي أنجز عملها هناك.	ترى وجوب العمل المقدس هناك.	هوللي شعرت عملها قد كان معمول.	شعر هوللي التي أنجزت عملها هناك.
	<i>She realized how that sounded.</i>	أدركت كيف أن بدا.	ادراكها كيف بدا.	أدركت كم الذي بدا.	أدركت كيف أن بدأ.
	<i>That did not sound right.</i>	يبدو أن لا حق.	لا صوت الحق.	لم الذي بصوت صحيح.	التي لا صوت الحق.

Table (2) Samples of Software Translators Output of Relative Non-Relative 'That' Clauses Input

The present software tools are unable to translate the word appropriately, nor can it connect the relative 'that' with its referent and in some instances, it cannot even tell whether 'that' is a relative word or a demonstrative. This highly signifies the kind of problems found in machine translation products, and motivates the present work. Our work results shall be compared to Google, the most successful translation software.

Translation Methodology of Google: Statistical Approach

Although Google translator provides translations among a large number of languages, the accuracies vary greatly. This is due to the fact that Google does not apply grammatical rules, since its algorithms are based on 'statistical' analysis rather than traditional 'rule-based' analysis. It is based on a method called statistical machine translation. Google has claimed that promising results were obtained using a statistical MT engine. Such engine (Vasta et al, 2010: 25) tries to extract what configuration of which words yields the most equivalent translation of a given input sentence. Hence, the translated text can often contain obvious errors. Such analysis can lead to many inaccurate results as it is clarified in the following samples translated by Google:

1 – A- She intended on beginning her day by collecting that torn envelope.

وقالت انها تهدف على بدايه يومها من خلال جمع المغلف الذي مزقته.

B – So you remember us having that conversation on the phone.

انت تتذكر لنا وجود تلك المحادثة على الهاتف.

In these examples, 'that' comes after a gerund and it should be translated either into [تلك or نلك] according to the word following it, i.e., (masculine or feminine). In 1- B, Google recognized that but in A, with the same structure it couldn't.

2 - A- We say that the earth is round.

نقول أن الارض كروية.

B - I'm sure that everything will work out.

انا متأكد من ان كل شيء سيكون العمل بها.

3 – A- Today I met the student that his father taught us last year.

اليوم التقيت الطالب ان والده علمتنا العام الماضي.

B – He is the only candidate that no one voted for.

انه هو المرشح الوحيد أن لا احد صوت له.

In 2, 'that' is not relative while in 3 it is relative. It can be noticed that in the previous examples Google couldn't differentiate between the two cases and it translated both of them into [ان] where in 3 both of them should be translated into [الذي]. Notice the following example:

4 – A- she intended on beginning her day by collecting that torn envelope.

وقالت انها تهدف على بداية يومها من خلال جمع المغلف الذي مزقته.

B – Holly felt that her work there had been done.

شعرت هوللي الذي انجز عملها هناك.

5 – The boy that she saw in the bus is my brother.

الصبي الذي رآته في الحافلة هو أخي.

In 4, A and B, 'that' is not relative and it should be translated into [لك]. In 5, it is relative and Google could not also make a difference and it translates both cases 4 and 5 with [الذي].

Translation Methodology of the Module: Rule-Based Approach

Rule-based approach is a transfer approach which operates in three stages (Hutchins, 1995: 432): The first stage converts SL text into abstract SL oriented representations (analysis). The second stage deals with lexical differences between languages and converts the representations resulted from the first stage into corresponding TL representations (transfer). Finally the third stage generates the TL text (generation).

Transfer systems require three types of dictionaries for the SL and TL. Those contain morphological, grammatical, and semantic information. Such components approve what is needed for a machine to achieve a higher degree outcome.

Here are some of the rules proposed for the module to give a satisfying outcome which is better than that used by Google:

<p>if the head noun is S M. الذي - 'That' is translated into if the head noun is S F or INA P F. التي - 'That' is translated into if the head noun is Dual M. الَّذَانِ - 'That' is translated into</p> <p>ذلك 'That' is translated into when it is followed by M N. when it is followed by F N / F NP. تلك - 'That' is translated into</p> <p>أُنْ 'That' is translated into - When it is preceded by Proper Name and followed by a Pronoun.</p>

Table (3) Translation Rules Examples

Designed Architecture of the Module:

The Electronic Dictionary

The translation of the input sentence cannot be done without certain data included in the dictionary. When the sentence is entered into the module, 'that' is translated according to the

word that precedes or follows it. These words have certain features integrated in the dictionary.

The following describes the structure of entries for the proposed dictionary for the relative *'that'*:

- Word (head noun) contains: word category (Noun, Adj, Adv, Pronoun...etc), Animacy (Animate/ Inanimate), definiteness (Definite/ Indefinite), gender (Masculine/ Feminine) and number (Singular/ Plural/ Dual), e.g.:
 - ("success", "M", "S", "N", "INA").
 - ("you", "M", "S", "PRO", "A").
 - ("Ahmed", "M", "S", "PN", "A").
 - ("boy", "M", "S", "N", "A").
 - ("much", "M", "S", "N", "INA").
 - ("few", "F", "S", "Q", "A").
 - ("music", "F", "S", "N", "INA").

The translation of relative *'that'* is determined according to these entries, e.g., a sentence like *'The students that passed the exam were so happy'*, the head noun *'students'* is preceded by the definite article *'the'* and it is a plural, masculine noun so the translation would be *الذين*.

As to the translation of non-relative *'that'*, the word that precedes *'that'* or follows it has the following structure of entries (either both or one of them):

- Word category.
- Gender (Masculine/ Feminine), e.g.:
 - ("accident", "M", "N").
 - ("did", "V").
 - ("night", "F", "N").
 - ("much", "Q").
 - ("and", "CON").

For instance, in a sentence like *'Since that accident I did not talk to her'*, the word *'accident'* is a masculine noun so the translation would be *ذلك*.

Sentence Test

The entire translation process is performed in steps. First, the input sentence is 'preprocessed' with various rules. Then, in the 'matching' step, the module searches the appropriate translation rule of the input sentence. The third step is 'transfer', which is the generation of the translation of each matched rule and on the last step; the module generates the complete Arabic translation of the word '*that*'. The following three steps explain the whole process:

1 Preprocessing

The sentence is inserted in one input. The module tests the sentence to see whether it is relative or not by choosing the appropriate rule for the source sentence under which it should be decided whether the word '*that*' is relative or not. Each sentence is submitted to one of the set of rules built inside the module. (See table 3) If the sentence is relative and matches one of the rules which are specified for the relative case; the head N is also checked by the dictionary entries which are specified for that head N to see if it is definite or indefinite.

2 Matching

In the matching step, the Arabic translation matches the preprocessed input sentence. If the head noun is definite, the sentence is translated by matching the dictionary entries which are specified for that head noun. If the head noun is indefinite '*that*' is going to take non translation 'zero translation' except when the head noun is [all/many/some, nobody], '*that*' is translated into [ما, ممن, ممن, ممن] respectively. If the sentence is not relative, it is applied to one of the rules specified for the non-relative case.

3 Transfer (Generation)

The output of the transfer step is the matching translations found in the previous step. The Engine now in this step applies

all the English rules to translate 'that'. The first step here is to find translations of 'that' in the dictionary. The dictionary relates the primitive form of the 'that' to the appropriate primitive form of the Arabic 'that'. The whole structure of the module is shown in Figure (1).

For example, when 'that' generates the translation as 'لك' it first checks the dictionary and calls the rule which says that the head N must be masculine and singular and so on:

trans ("لك):- ("M", "N"), !.

trans ("تلك):- ("F", "N"), !.

Trans ("الذي):- ("M", "S"), !.

trans ("التي):- ("F", "S"), !.

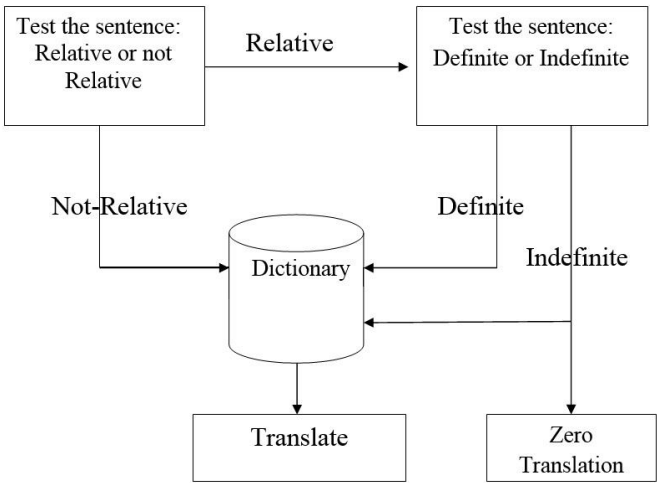


Figure (1)
The Program Structure

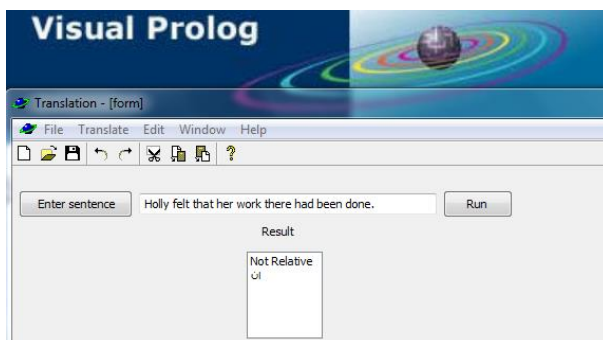
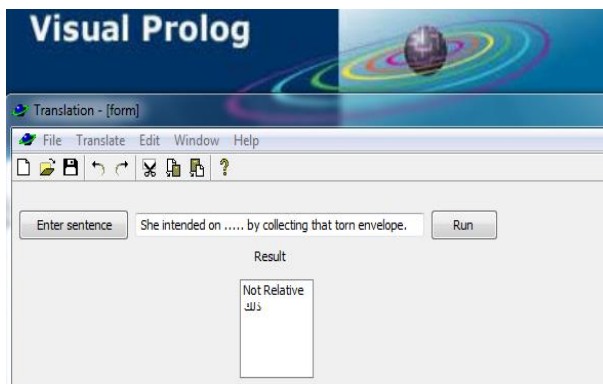
4 Module Results

The module presents the English word 'that' and its equivalent translation in Arabic. The Module was tested in more than 100

sentences. The following are some examples of the module results compared to Google and human translation:

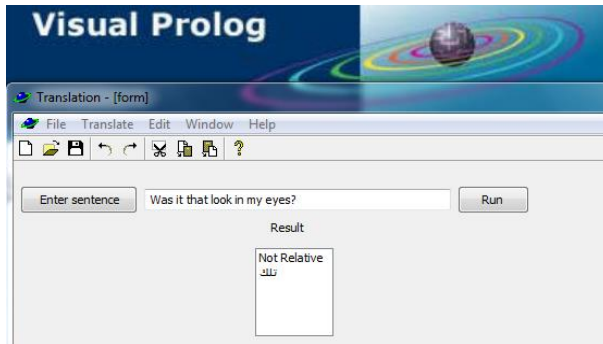
1 – 'That' is translated into 'ذلك' when it is followed by a masculine N or masculine NP and it is translated into 'ان' when it is preceded by a pronoun + V:

Sentence	Google Translation	Human Translation	Module Translation
<i>She intended by beginning her day by collecting that <u>torn envelope</u>.</i>	و قالت انها تهدف على بداية يومها من خلال جمع <u>المغلف الذي مزقته</u> .	ذلك المغلف الممزق	ذلك
<i>Holly felt that her work there had been done.</i>	شعرت هوللي الذي انجز عملها هناك.	ان	ان



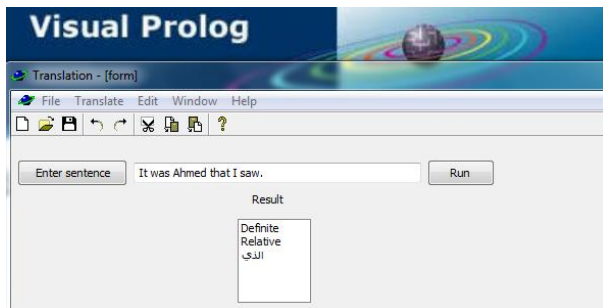
2 'That' is translated into 'تلك' when it is followed by a feminine N or feminine NP:

Sentence	Google Translation	Human Translation	Module Translation
Was it <i>that</i> look in my eyes?	وكان إن نظرت في عيني؟	تلك النظرة	تلك



3 – 'That' is translated into 'الذي' when the head N is: singular masculine N or Proper Name:

Sentence	Google Translation	Human Translation	Module Translation
It was <i>Ahmed</i> that I saw.	كان احمد ان رأيت.	احمد الذي	الذي



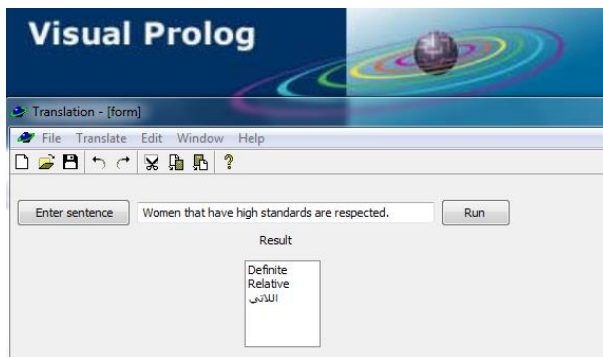
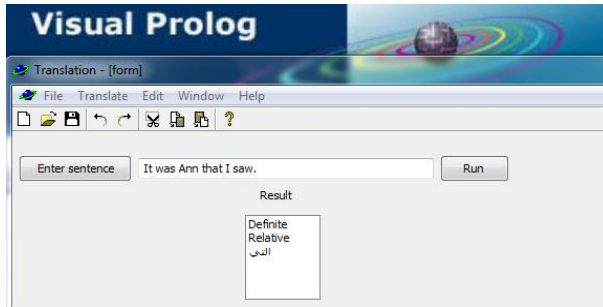
4– 'That' is translated into 'التي' when the head N is: singular feminine [N or Pronoun or Proper Name] and into 'اللاتي' when the head N is animate plural feminine:

Sentence	Google Translation	Human Translation	Module Translation
It was <i>Ann</i> that I saw.	إن كان الذي رأيت.	آن التي	التي
<i>Women</i> that have high	المرأة التي لديها مستويات	النساء اللاتي	اللاتي

Farah Abbas Abo Al Timen- **Module for Machine Translating Relative and Non-Relative "That" from English into Arabic**

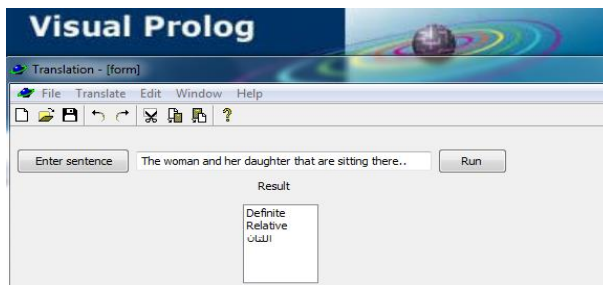
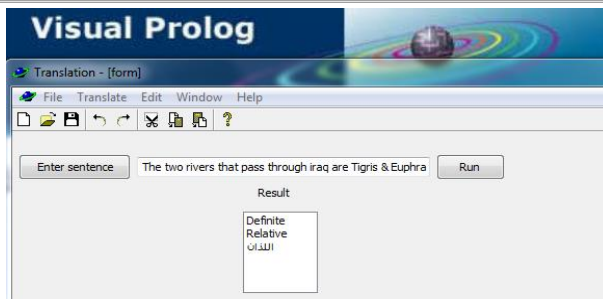
standards are respected.

عالية مما الاحترام.



5 – 'That' is translated into '**الَّذان**' when the head N is: masculine dual and '**اللتان**' when the head N is feminine dual:

Sentence	Google Translation	Human Translation	Module Translation
<i>The <u>two rivers</u> that pass through Iraq are Tigris and Euphrates.</i>	<i>النهرين التي تمر عبر العراق دجلة والفرات.</i>	<i>النهران اللذان</i>	<i>الَّذان</i>
<i>The <u>woman and her daughter</u> that are sitting there are my neighbors.</i>	<i>المرأة و ابنتها التي تجلس هناك جيراني.</i>	<i>المرأة وابنتها اللتان</i>	<i>اللتان</i>



REFERENCES

- Aljlayl, M., Frieder, and D. Grossman (2004). *On Arabic-English Cross Language Information Retrieval: A Machine Translation Approach*.
- Arnold, Douglas, Lorna Balkan, Siety Meijer, R. Lee Humphreys and Louisa Sadler (1994). *Machine Translation: An Introductory Guide*. Blackwell; London.
- Attia, Mohammed Abdel-Aal (2002). *Implication in the Agreement Features in Machine Translation*. MA thesis, Al Azhar University.
- Hutchins, W. John (1995). *Trends in machine translation research*. Unpublished. Written 1995 for Sprog och Erhverv.
- <http://www.hutchinsweb.me.uk/SprogErhverv-1995.pdf>
[Accessed: 11, Mar, 2012].
- Vasta, Mukesh, Nikita Joshi, and Sumit Goswami (2010). *Statistical Machine Translation*. DESIDOC Journal of

Library & Information Technology. Vol. 30, No. 4, pp. 25-32.

Wilks, Yorick (2009). *Machine Translation: Its Scope and Limits*. Springer; New York.

Notes:

1 Google Software link: <http://translate.google.com> [Accessed: 23-10-2012].

2 Babylon Software translation link: <http://translation.babylon.com/english/to-arabic> [Accessed: 24-11-2012].

3 Golden Al-Wafi Software link: <http://www.down32.com/software/portable-apps/129-golden-al-wafi-translator.html> [Accessed: 31-12-2012].

4 Bing Software Link: <http://www.bing.com/translator> [Accessed: 28-12-2012].